Guidelines for the Identification of a Stochastic Model for the Genetic Toggle Switch*

Brian Munsky[1,*] and Mustafa Khammash[2]

[1]Center for Nonlinear Studies and
Computer, Computational and Statistical Sciences Division,
Los Alamos National Laboratory,
Los Alamos, NM 87545 USA
*To whom correspondence should be addressed:
E-mail: brian.munsky@gmail.com

[2]Center for Control, Dynamical Systems and Computations and
The Department of Mechanical Engineering,
University of California,
Santa Barbara, CA 93106, USA.

May 28, 2010

Abstract

Due to the inherently random and discrete nature of genes, RNAs, and proteins within living cells, there can be a wide range of variability both over time in a single cell and from cell to cell in a population of genetically identical cells. Different mechanisms and reaction rates help shape this variability in different ways, and the resulting cell-to-cell variability can be quantitatively measured using techniques such as time-lapse microscopy and fluorescence activated cell sorting (or flow cytometry). It has been shown that these measurements can help to constrain the parameters and mechanisms of stochastic gene regulatory models. In this work, finite state projection approaches are used to explore the possibility of identifying the parameters of a specific stochastic model for the genetic toggle switch consisting of mutually inhibiting proteins: LacI and λcI [1, 2]. This article explores the possibility of identifying the model parameters from different types of statistical information, such as mean expression levels, LacI protein distributions and LacI-λcI multivariate distributions. It is determined that although the toggle model parameters cannot be uniquely identified from measurements that track just the LacI variability, the parameters could be identified from measurements of the cell-to-cell variability in both regulatory proteins. Based upon the simulated data and the computational investigations of this study, experiments are proposed that could enable this identification.

1 Introduction

A key issue facing modelers of gene regulatory systems is that rare chemical components (e.g., genes, RNA molecules, and proteins) can lead to large amounts cellular variability [3, 4, 5, 6, 7, 8, 9]. This variability has attracted much recent attention, and it is well established that different system mechanisms will affect variability in different manners. In some systems, variability enhances dynamic signals via stochastic focussing [10]; in other cases it may cause or enhance resonant fluctuations [11]; some other network topologies may result in stochastic switching [12, 13, 14, 15]; and in many systems deleterious variability may be repressed [16].

There are a number of well-established experimental techniques in which fluorescent markers are used to highlight the molecular variability of single cells[17]. The most common of these techniques is flow cytometry [18], which enables researchers to perform hundreds of millions of controlled single cell experiments all at the same time. For this technique, cell strains can be engineered to express fluorescent proteins such as green, yellow or cyan fluorescent protein (*gfp, yfp,cfp*), all of which can be measured in a fraction of a second for each cell. Different cultures of the same cell strain can be perturbed with different inputs, at different levels, and at different times. A researcher can measure a million cells in a minute and test forty or more different conditions in a single hour. And, unlike population level and/or *in vitro* techniques for measuring biological parameters (gels, immunoblotting, etc...), the cells can be grown naturally throughout the experiment, and measurements are made on individual living cells.

The more recently discovered technique of single molecule fluorescence *in situ* hybridization (FISH, [19]) allows researchers to tag and count specific mRNA populations in individual fixed cells. With automated imaging techniques it is possible to count mRNA molecules in thousands of individual cells, thereby obtaining precise distributions under carefully controlled experimental conditions. Flow cytometry and FISH based single cell microscopy are highly complementary approaches to study variability. Whereas cytometry with fluorescent proteins can measure post-translational activity, single cell microscopy with FISH mRNA markings can measure pre-translational regulation. Although flow cytometry allows for the measurement of more individual cells, single cell microscopy allows for more precise measurements of individual cells.

Using either or both experimental approaches, the measured variability in gene expression offers a surprisingly rich source of information about system parameters and dynamics. This vast amount of data is often far more complicated to analyze, and in many computational studies, cell-to-cell variability has been viewed as a computational nuisance–a viewpoint that isn't helped by the term "noise," which is frequently attached to this variability. If one does not include intrinsic variability in a model, then one cannot hope to capture certain cellular behaviors. However, in many cases, the inclusion of model stochasticity results in an explosion of computational complexity. For many other cases, the intrinsic variability of gene regulation can be analyzed techniques such as kinetic Monte Carlo algorithms [20] and stochastic differential equation approaches [21]. In turn, these KMC approaches have been improved with various sampling approaches [22, 23], τ leaping approaches [24], and time scale separation schemes [25, 26]. In a different direction, many researchers have sought to directly solve for the evolution of probability distributions using various techniques such as linear noise approximations [27, 28], moment closure [29, 30] and matching [31] techniques, moment generating functions [32], spectral methods [33] and finite state projection approaches [34, 35, 36, 37]. At present, none of these approaches suffices to handle all systems, and

there remains much work to be done to improve our computational capabilities. However, as these tools develop, it becomes more possible to overcome the obstacle of intrinsic noise and gain significant benefits from analytical studies.

When it is possible to overcome the obstacles presented by cellular variability, the inclusion of system stochasticity can reveal a wealth of additional information regarding the dynamics of biochemical networks. Analyses of how variability is affected by different regulatory architectures provide a new tool with which to compare and contrast different possibilities for evolutionary design [38]. Alternatively, when this design is not known *a priori*, analyses of cellular variability may help to discover it [39, 40, 41]. For example, different logical structures such as AND or OR gates can be discovered in two component regulatory systems by examining the stationary transmission of the cell variability through the network [39], or correlations of different aspects of cell expression at many time points can reveal different causal relationships between genes within a network [40]. Similarly, measuring and analyzing the statistics of gene regulatory responses in certain conditions can help to identify system parameters and develop quantitative, predictive models for certain systems [42, 43].

In this article, we use computational analyses to demonstrate the usefulness of combining stochastic analyses and single cell measurements to identify gene regulatory models. We follow a similar approach to that in [42] in that we will apply finite state projection (FSP–[34, 36, 37]) tools to conduct stochastic analyses and parameter identification for the genetic toggle switch from [1, 2]. In the following section, we provide a brief background on previous analyses of the toggle switch, and we discuss the application of FSP analysis tools to this system. Then, in Section 3 we use these tools and simulated data to determine the types of additional experiments that are necessary to fully identify the parameters of the toggle model. Finally, in Section 4 we summarize our findings and make a few concluding remarks.

2 Background

The toggle switch, composed of the mutually inhibiting genes *lacI* and λcI, was first constructed and presented in [1] and then extended in [2] to be used as a sensor of environmental influences, such as radiation or external chemical signals. This switch is a construction of two genes, each of whose protein, λcI or LacI, inhibits the production of the other (see Fig. 1A). With exposure to Ultraviolet light (UV) or mitomycin C (MMC), the SOS pathway results in RecA coproteases, which increase the degradation rate of λcI. As a result, different amounts of UV or MMC change the trade-off between λcI and LacI molecules. The output of the mechanism is GFP, which is also controlled by the same promoter as *lacI* and is assumed to be expressed at the same level as LacI. Depending upon environmental conditions, the system exhibits a bias toward one phenotype or another (i.e., it either has a high level of LacI and GFP and is highly fluorescent or it has a high level of λcI and is not fluorescent). A vast number of models have been proposed to describe this and other toggle switches, including deterministic [2] and stochastic versions [44, 15, 45] to name a few. This study considers a particular stochastic model similar to that in [15] and aims to determine the types of experimental data necessary to determine the model parameters. To explore these requirements, we use simulated data of the type that can be measured using flow cytometry experiments such as those conducted in [2].

The stochastic model of the toggle switch is composed of four non-linear production / degradation reactions given by:

$$\begin{array}{llllllll} \mathrm{R}_1 & ; & \mathrm{R}_2 & ; & \mathrm{R}_3 & ; & \mathrm{R}_4 \\ \emptyset \rightarrow \lambda\mathrm{cI} & ; & \lambda\mathrm{cI} \rightarrow \emptyset & ; & \emptyset \rightarrow \mathrm{LacI} & ; & \mathrm{LacI} \rightarrow \emptyset. \end{array} \tag{1}$$

The rates of these reactions, $\mathbf{w}(\lambda\mathrm{cI}, \mathrm{LacI}, \mathbf{\Lambda}) = [w_1(\lambda\mathrm{cI}, \mathrm{LacI}, \mathbf{\Lambda}), \ldots, w_4(\lambda\mathrm{cI}, \mathrm{LacI}, \mathbf{\Lambda})]$ depend upon the populations of the proteins, λcI and LacI, as well as the set of non-negative parameters, $\{k^{(0,1)}_{\lambda\mathrm{cI}}, \alpha_{\mathrm{LacI}}, \eta_{\mathrm{LacI}}, k^{(0,1)}_{\mathrm{LacI}}, \alpha_{\lambda\mathrm{cI}}, \eta_{\lambda\mathrm{cI}}, \delta_{\mathrm{LacI}}, \delta_{\lambda\mathrm{cI}}\}$, according to:

$$w_1 = k^{(0)}_{\lambda\mathrm{cI}} + \frac{k^{(1)}_{\lambda\mathrm{cI}}}{1 + \alpha_{LacI}[\mathrm{LacI}]^{\eta_{\mathrm{LacI}}}}; \qquad w_2 = \delta_{\lambda\mathrm{cI}}[\lambda\mathrm{cI}];$$

$$w_3 = k^{(0)}_{\mathrm{LacI}} + \frac{k^{(1)}_{\mathrm{LacI}}}{1 + \alpha_{\lambda\mathrm{cI}}[\lambda\mathrm{cI}]^{\eta_{\lambda\mathrm{cI}}}}; \qquad w_4 = \delta_{\mathrm{LacI}}[\mathrm{LacI}],$$

In the model, the λcI degradation parameter, $\delta_{\lambda\mathrm{cI}}$, takes on different values depending upon environmental influences such as UV radiation, while the remaining nine parameters are assumed to be independent of environmental conditions. Thus, the full parameters set is represented by:

$$\mathbf{\Lambda} = \{k_{\lambda\mathrm{cI}}^{(0,1)}, \alpha_{\mathrm{LacI}}, \eta_{\mathrm{LacI}}, k_{\mathrm{LacI}}^{(0,1)}, \alpha_{\lambda\mathrm{cI}}, \eta_{\lambda\mathrm{cI}}, \delta_{\mathrm{LacI}}\}, \{\delta_{\lambda\mathrm{cI}}(0), \delta_{\lambda\mathrm{cI}}(6), \delta_{\lambda\mathrm{cI}}(12)\} \in \{\mathbb{R}^9_{\geq 0}, \mathbb{R}^3_{\geq 0}\}.$$

At present there is insufficient evidence to define every parameter of the toggle switch model, and the current study relies upon simulated data in an effort to determine what additional experiments are necessary. To generate this data, we have chosen a reference parameter set as follows:

$$\begin{array}{lll} k_{\lambda\mathrm{cI}}^{(0)} = 6.8\times10^{-5}\ \mathrm{s}^{-1} & k_{\lambda\mathrm{cI}}^{(1)} = 1.6\times10^{-2}\ \mathrm{s}^{-1} & \alpha_{\mathrm{LacI}} = 6.1\times10^{-3}\ \mathrm{N}^{-\eta_{\mathrm{LacI}}} \\ k_{\mathrm{LacI}}^{(0)} = 2.2\times10^{-3}\ \mathrm{s}^{-1} & k_{\mathrm{LacI}}^{(1)} = 1.7\times10^{-2}\ \mathrm{s}^{-1} & \alpha_{\lambda\mathrm{cI}} = 2.6\times10^{-3}\ \mathrm{N}^{-\eta_{\lambda\mathrm{cI}}} \\ \eta_{\mathrm{LacI}} = 2.1\times10^{-0} & \eta_{\lambda\mathrm{cI}} = 3.0\times10^{-0} & \delta_{\mathrm{LacI}} = 3.8\times10^{-4}\ \mathrm{N}^{-1}\mathrm{s}^{-1} \\ \delta_{\lambda\mathrm{cI}}(0) = 3.8\times10^{-4}\ \mathrm{N}^{-1}\mathrm{s}^{-1} & \delta_{\lambda\mathrm{cI}}(6) = 6.7\times10^{-4}\ \mathrm{N}^{-1}\mathrm{s}^{-1} & \delta_{\lambda\mathrm{cI}}(12) = 1.5\times10^{-3}\ \mathrm{N}^{-1}\mathrm{s}^{-1}, \end{array} \tag{2}$$

where the notation N corresponds to the integer number of molecules of the relevant reacting species. These parameters have been chosen partially to match values in the literature and partially to match the qualitative behaviors of the measured fluorescence histograms from [2] (i.e. the model expresses a small amount of LacI at low UV levels, a large amount of LacI at higher UV levels and has a bimodal distribution at intermediate UV levels–see Fig. 1). In the context of the current study, the most important of these parameters are the half-lives of LacI and λcI, which effectively set the timescale of the system's transient responses. Since both proteins are extremely stable and have half-lives that are much longer than the cell division time, $\delta_{\lambda\mathrm{cI}}(0)$ and $\delta_{\mathrm{LacI}}(0)$ have both been set to match a dilution half live of 30 minutes (cell division time). The cooperativity of λcI binding, $\eta_{\lambda\mathrm{cI}}$ is chosen to be 3 to reflect the three binding sites of λcI to the $\mathrm{P_L}$ promoter. The cooperativity of LacI binding, η_{LacI} is set to 2.1 as identified in [42]. The remaining parameters have been chosen to match the qualitative behavior of the system as measured in [2] at the different levels of UV radiation.

At this point, no attempt has yet been made to fit the quantitative values of these parameters for the current model. For such a fit to provide much insight, one would first need to calibrate for background fluorescence and allow for the extrinsic variability in gfp fluorescence as explored in [42]. This would require far more experimental data than is currently available for this system. With the use of 96- and 384-well plate auto-samplers, today's flow cytometry equipment is capable of providing this additional information for little additional experimental cost, and the acquisition of this data will not be a limiting step in the identification process.

2.1 Stochastic Analysis of the Toggle Switch

By assuming that cells constitute a well-mixed environment, one can model biochemical populations with a jump-Markov process. As mentioned above, there are numerous approaches to analyze such processes, and many researchers are actively involved in developing new approaches. For the purpose of this examination, the choice of method does not matter, so long as it provides an accurate and efficient solution. Because the finite state projection (FSP -[34, 35, 46, 36, 37]) approach provides a direct accuracy guarantee on the solution of the master equation, it is a natural choice with which to conduct this investigation. In this subsection, we provide a brief background on the use of the FSP approach for the modeling of the toggle switch.

Under the reactions presented in the preceding section, the joint probability density of having i molecules of LacI and j molecules of λcI evolves according the set of linear ordinary differential equations,

$$\begin{aligned} \frac{dp_{i,j}(t,\mathbf{\Lambda})}{dt} = &-\sum_{k=1}^{4} w_k(i,j,\mathbf{\Lambda})p_{i,j}(t,\mathbf{\Lambda}) \\ &+ w_1(j,\mathbf{\Lambda})p_{i-1,j}(t,\mathbf{\Lambda}) + w_2(i+1,\mathbf{\Lambda})p_{i+1,j}(t,\mathbf{\Lambda}) \\ &+ w_3(i,\mathbf{\Lambda})p_{i,j-1}(t,\mathbf{\Lambda}) + w_4(j+1,\mathbf{\Lambda})p_{i,j+1}(t,\mathbf{\Lambda}), \end{aligned}$$

which is typically known as the (chemical) Master Equation [27]. Because the toggle reaction rates, w_1 and w_2, are non-linear, the ME has no known closed form solution, but the finite state projection approach [34, 36] allows us to approximate the solution to any pre-specified degree of accuracy. Given the infinitesimal generator matrix, $\mathbf{A}(\mathbf{\Lambda})$, the initial probability distribution, $\mathbf{P}(0)$, and a chosen error tolerance, $\varepsilon > 0$, we can systematically find a finite projection system $\dot{\mathbf{P}}^{FSP}(t,\mathbf{\Lambda}) = \mathbf{A}_J(\mathbf{\Lambda}) \cdot \mathbf{P}^{FSP}(t,\mathbf{\Lambda})$ such that:

$$\left\| \begin{bmatrix} \mathbf{P}_J(t,\mathbf{\Lambda}) \\ \mathbf{P}_{J'}(t,\mathbf{\Lambda}) \end{bmatrix} - \begin{bmatrix} \mathbf{P}^{FSP}(t,\mathbf{\Lambda}) \\ \mathbf{0} \end{bmatrix} \right\|_1 \leq \varepsilon, \text{ and } \mathbf{P}^{FSP}(0) = \mathbf{P}_J(0), \tag{3}$$

where the index vector J denotes the set of states included in the projection, $\mathbf{P}_J$ is the corresponding probabilities of those states, and $\mathbf{A}_J$ is the corresponding principle submatrix of $\mathbf{A}$ [34, 36].

For this study, we have chosen the projection set, J, to include all states that satisfy: $i \leq N_1$, $j \leq N_2$, and $(i-3)(j-3)^2 \leq N_3$, where the values of N_1, N_2, and N_3 are allowed to increase until the projection error, ε, is less than 10^{-6}. The actual expansion is carried out as described in [46], where the values of N_1, N_2 and N_3 are increased depending upon the dominant directions of probability leakage out of the current projection space (see Section 2.1 of [46] for more details). For the reference parameter set the final projection space is defined by the triplet $(N_1, N_2, N_3) = (90, 90, 18400)$ for the case with $0J/m^2$ UV radiation, $(N_1, N_2, N_3) = (50, 110, 17600)$ for the case with $6J/m^2$ UV radiation, and $(N_1, N_2, N_3) = (30, 110, 12800)$ for the case with $12J/m^2$ UV radiation. Other parameter sets require different projection spaces to reach the desired error tolerance, and the projection space is expanded and contracted during each parameter search.

2.2 Fitting the Toggle Model to Simulated Data

In most cases, the full joint probability distribution is not measured, and only an expected value or marginal distribution is to be considered. To explore how different quantities of information may lead to different identification results, we consider identification strategies using five such statistical quantities:

(1) Only the mean level of LacI, denoted as μ_{LacI}.
(2) The mean levels of LacI and λcI, denoted as $\mu_{\{\text{LacI},\lambda\text{cI}\}} := \{\mu_{\text{LacI}}, \mu_{\lambda\text{cI}}\}$.
(3) The marginal distribution of LacI, denoted as f_{LacI}.
(4) The marginal distributions of LacI and λcI, denoted as $f_{\{\text{LacI},\lambda\text{cI}\}} := \{f_{\text{LacI}}, f_{\lambda\text{cI}}\}$.
(5) The full joint distributions of LacI and λcI, denoted as Full $\mathbf{P}$.

Each of these is naturally represented as a linear projection of the full distribution. For example, the mean level of LacI and the marginal distribution of LacI are given by:

$$\mu_{\text{LacI}}(\mathbf{\Lambda}, t) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} i p_{i,j}(\mathbf{\Lambda}, t), \qquad \text{and} \qquad f_{\text{LacI}}(\mathbf{\Lambda}, t) = \sum_{j=0}^{\infty} p_{i,j}(\mathbf{\Lambda}, t)$$

Equivalently, one could write each of the statistical quantities (1)-(5) above in matrix form as $\mathbf{y}(\mathbf{\Lambda}, t) = \mathbf{C}\mathbf{P}(\mathbf{\Lambda}, t)$, where the output matrix $\mathbf{C}$ depends upon the statistical quantity to be considered. We note that if all of the reaction rates were affine linear, then the mean behaviors of μ_{LacI} and $\mu_{\lambda\text{cI}}$ would be equivalent to the deterministic mass action kinetics model of the same system. Because the toggle model has non-linear production terms, this equivalence does not hold. It is also important to note that the toggle switch gives rise to a relatively complicated bimodal distribution, which is not adequately be described with a small number of statistical moments, so low order moment based approaches such as those in [27, 28, 29, 30, 31] should be expected to yield incorrect results for this model.

With the FSP solution approach in hand, the identification procedure is relatively simple–we find the parameter arguments, $\mathbf{\Lambda}^\star$, that minimizes the difference between the measured statistical quantity, $\tilde{\mathbf{y}}(t)$, and the numerical solution of that quantity, $\mathbf{y}(\mathbf{\Lambda}, t)$:

$$\mathbf{\Lambda}^\star := \text{argmin}_{\mathbf{\Lambda}} \left|\tilde{\mathbf{y}}(t) - \mathbf{y}(t, \mathbf{\Lambda})\right|_p . \tag{4}$$

More specifically, because we will frequently be comparing probability distributions, we use the one-norm (i.e., $p = 1$) difference for a number of reasons: First, the FSP approach (Eqn. 3) directly computes the exact 1-norm error in the solution of the master equation, which then provides exact bounds on the 1-norm difference between a measured and predicted distributions. Second, the 1-norm of any probability distribution is exactly one and the 1-norm difference between two distributions lies between zero for a perfect match and two for a perfect mismatch. Finally, in our experience 1-norm optimizations provide distributions that better match the full qualitative shape of measured distributions, whereas other norms (such as the Euclidian norm) apply too much importance to peaks of these distributions, independent of how much probability measure is contained in those peaks.

In this study, we begin each minimization in (4) with the same initial guess:

$$\begin{array}{lll}
k^{(0)}_{\lambda\text{cI}} = 10^{-4}\ \text{s}^{-1} & k^{(1)}_{\lambda\text{cI}} = 10^{-2}\ \text{s}^{-1} & \alpha_{\text{LacI}} = 10^{-2}\ \text{N}^{-\eta_{\text{LacI}}} \\
k^{(0)}_{\text{LacI}} = 10^{-3}\ \text{s}^{-1} & k^{(1)}_{\text{LacI}} = 10^{-2}\ \text{s}^{-1} & \alpha_{\lambda\text{cI}} = 10^{-3}\ \text{N}^{-\eta_{\lambda\text{cI}}} \\
\eta_{\text{LacI}} = 3 & \eta_{\lambda\text{cI}} = 3 & \delta_{\text{LacI}} = 10^{-4}\ \text{N}^{-1}\text{s}^{-1} \\
\delta_{\lambda\text{cI}}(0) = 10^{-4}\ \text{N}^{-1}\text{s}^{-1} & \delta_{\lambda\text{cI}}(6) = 10^{-3}\ \text{N}^{-1}\text{s}^{-1} & \delta_{\lambda\text{cI}}(12) = 10^{-3}\ \text{N}^{-1}\text{s}^{-1}.
\end{array}$$

and then update this guess with iterations of gradient-based and simulated annealing searches until the objective function has decreased by at least two orders of magnitude. In the numerical studies presented here, it is known *a priori* that the "true" parameter set gives an exact match and satisfies this criteria for any p-norm. If the

optimization terminates at a much different parameter set, then the identification is not unique for the data set $\tilde{\mathbf{y}}$. In more realistic practice, local minima must be discarded by using multiple, randomly generated initial parameter guesses. In that case, an optimal parameter set would be considered to be unique if the given solution yields the smallest achieved value for the objective function and if that parameter has been repeatedly found during independent optimization procedures.

3 Results

The model in the previous section and the parameters in (2) have been used to numerically generate the joint probability distribution of the numbers of LacI and λcI molecules. We have assumed an initial condition of zero molecules of each species, and have solved the master equation for the distribution at times of {1,2,3,4,5} hours later. In this section, we consider each of the five identification strategies listed above and examine how successful each approach may be for this identification. For each strategy, we first consider the identification using the relevant data at 5hrs only and later with the same information at all five time points. Each identification strategy results in a parameter set that captures the relevant data to within a one norm difference of 10^{-1}, and Tables 1 and 2 list the relative values of these identified parameters. In most cases, the identification is insufficiently constrained, and the identified parameter set is not unique (i.e., it is different than the original parameter set). For these, the parameter set may match some portions of the simulated data, but not others. Table 3 tabulates these differences for the ten different parameter sets. We note that the numbers in Tables 1-3 should be viewed as qualitative results as the identified parameters sets {1-4,6,7} depend upon the initial guess and are not unique.

3.1 Identification from a single time point.

Table 1 lists the relative values of the identified parameters compared to their actual values (Reference set #0). These parameters have been identified from different statistical quantities taken from the simulated data at a time of 1 hour. The following subsections discuss the corresponding success or failures in these identification attempts.

3.1.1 Identification using mean level of LacI at $t = 5hr$.

In many modeling endeavors, researchers concentrate their efforts solely on matching the mean behaviors of the observable systems. In other situations, there are often good reasons for this choice. First, most experimental measurements are taken at the population level using lysed cells, and data on the cell-to-cell variability is not available. Second, if one is only interested in the mean level behavior, then it is often possible to utilize deterministic models, which are typically more computationally tractable. These reasons are far less compelling in this situation. When all measurements are taken using flow cytometry or another single cell technique, data on cell-to-cell variability is readily available. It would be a waste to ignore this information. Furthermore, due to strong nonlinearities in the LacI and λcI production rates, the corresponding deterministic model fails to capture the mean behavior of the true stochastic system (results not shown).

We first explore if it is possible to identify the unknown parameters from the mean LacI level, μ_{LacI}, at all three UV radiations levels and at a time of 5 hours. While it is not difficult to find a set of 12 parameters that match the mean level of LacI under the different levels of ultraviolet radiation, these parameters are not unique. Indeed, we find that parameter set #1 in Table 1 matches the mean LacI expression levels as can be seen by the blue bars in Fig. 2A (color online). However, the same parameter set gives a very poor prediction for the mean level of λcI as shown in Fig. 2B and Table 3. Thus, it is clear that the mean level of lacI is insufficient to identify the model parameters.

3.1.2 Identification using mean levels of LacI and λcI at $t = 5hr$.

We next consider the possibility of identifying all twelve parameters from the mean level of both LacI and λcI for all three UV levels and at a time of five hours after induction. Once again it is easy to find that parameter set #2 in Table 1 matches these mean levels for both species, as can be seen with the green bars plotted in Figs. 2A,B. On the other hand, closer inspection of the marginal probability distributions of LacI and λcI in (shown in green lines in Fig. 3A-F) reveals that this parameter set provides a poor prediction for the marginal distributions of the two chemical species, especially at low levels of UV induction (see Figs. 3{A,B,D,E} and Table 3).

3.1.3 Identification using marginal distribution of LacI at $t = 5hr$.

As discussed in [42], if one utilizes information of the cell-to-cell variability, then it become possible to better distinguish between models and parameter sets. With this in mind, we have attempted to conduct the identification from the marginal distribution of LacI at five hours after induction with the different levels of UV radiation. Parameter set #3 in Table 1 has been found to match this marginal distribution. However, we once again discover that the identification is not unique, which suggests that data regarding the variability of a single protein (i.e., the kind of data taken in [2]) is not sufficiently rich for complete identifiability. On the other hand, closer inspection of Table 2 reveals a fixed ratio of about 0.6 between many of the parameters from set #3 and the true parameter set. This suggests that the identification has drastically narrowed the space of possible parameter sets, but more information is needed to provide a unique parameter set for the proposed model. For example, we see that parameter set #3 gives poor prediction for the mean and marginal distribution for λcI as can be seen in Fig. 2B and Figs. 3{D-F} and Table 3. Thus, any additional measurements of λcI would help to further constrain the model.

3.1.4 Identification using marginal distribution of both LacI and λcI at $t = 5hr$.

Parameter set #4 has been identified from the simulated data using the marginal distributions of both species at all UV radiation levels and at the time of five hours after induction. Now the identified set correctly matches the mean and marginal distributions of LacI and λcI as shown in Figs. 2 and 3. However, there still remains a slight discrepancy between the full joint distribution as computed with the reference parameter set and the same joint distribution as computed with parameter set #4 (see Table 3).

3.1.5 Identification using the full joint LacI/λcI distribution at $t = 5hr$.

Table 3 shows that the joint distribution predicted by parameter sets #1-4 are different than the joint distribution found with the reference parameter set, suggesting that use of this full distribution would go a long way to further constraining the model. In our final attempt to identify the full parameter set from a single time point, we have conducted the identification based upon the full joint distribution of LacI and λcI at the time of 5 hours. In this case, it appears that identification of the parameters is indeed possible, and all of the parameters are identified within a very small distance of their actual values. As an aside, it is interesting to note that the optimization problem that uses the full distribution is much more easily solved than that which utilized only the marginal distributions. In these studies the identification based upon the whole distribution took less than an hour and converged to a much lower value than that based upon marginal distributions, which took several days. A full exploration of how these different data sets lead to better conditioning of the optimization procedure is left for future work.

3.2 Identification from multiple time points.

From above and from [42], it is clear that more statistical information provides a better chance for a system's identifiability. Similarly, it is equally important to conduct that identification using different time points, preferably during the system's transient response [42]. In the preceding subsection, we have attempted the identification based upon a single time point that occurred 5 hours after induction. We note that the degradation rates of LacI and λcI in the absence of UV radiation correspond to a half life of 30 minutes (i.e., the cell division time) and even faster with UV radiation. Thus, since the identification is being conducted after ten such half lives, we are likely missing much of the system's transient dynamics. We have therefore redone the identification using measurements at each hour after induction.

Table 2 lists the relative values of the identified parameters as found using this more detailed information. For a comparison of the system dynamics with the different parameter values, Fig. 4 shows the responses of the mean LacI and λcI levels. From the figure, it is clear that Parameter Set #6 matches the response of LacI over time (Panel A), but not that of λ_{cI} (squares in Panels D-F). Furthermore, Fig. 4 and Table 3 show that Parameter Set #7 matches the mean dynamics of both proteins, but is insufficient to complete the identification. On the other hand, we note that with the additional time points, the marginal distribution of LacI (Parameter Set #8) does a much better job of obtaining the correct parameter values for the parameters. Finally, even when the full distribution is used to identify the parameters, the use of multiple time points considerably improves the

approach–yielding more precise results and making the identification much less sensitive to unbiased measurement errors.

4 Conclusions

Many gene regulatory systems are characterized by small numbers of regulatory molecules, and therefore result in a large amount of variability both over time for single cells and from cell to cell in a clonal population. This variability is measurable with a number of experimental approaches, including time lapse fluorescence microscopy which can record the fluctuations of individual living cells and flow cytometry which allows for the rapid measurement of large populations of individual cells. To fully access the information available from these measurement techniques, it is necessary to utilize quantitative stochastic models which can capture these behaviors. Finite state projection (FSP–[34]) type solutions are ideal for these types of analyses because they specifically compute the transient dynamics of cell-to-cell probability distributions–precisely the types of information measurable with flow cytometry. By combining flow cytometry measurements with FSP approaches, it becomes easier to constrain gene regulatory models and even completely identify models and parameters for natural and synthetic circuits [42].

Similarly, computational analyses of stochastic gene regulatory models can help researchers to determine the types of experiments necessary to provide better understanding of a given system. In this work we have used FSP-based computational studies to explore the possibility of identifying parameters for a particular model of the genetic toggle switch [1, 2]. Our numerical studies have revealed that different identification strategies will produce varying degrees of success. As summarized in Table 3, each successive addition of more statistical information can help to further constrain the space of allowable parameters–eventually leading to a single unique point for this model. For example, parameters can be found to match the mean (or marginal distribution) of one protein but not the other. Similarly, models can be found that match the mean behaviors of the system, but which do not match the marginal distributions of one or more species. Finally, models can be found to match the marginal distributions of both species, but which do not match the full joint distribution of both species. However, for this particular model, only one set of parameters is capable of matching the full joint distribution at the time of 5 hours. Similarly, if one uses more experimental time points, then this also can further constrain the model and lead one to arrive at better models and more useful sets of parameters. In particular, we predict that if one could measure both LacI and λcI populations in large numbers of individual cells at a time resolution of about an hour, then one could in principle fully identify the parameters for the proposed model of the gene toggle switch. We note that models of greater complexity such as those including additional parameters related to transcription, translation, oligomerizaytion, and other kinetics will require more measurements for complete identification.

The necessary experimental measurements for the identification of the current model could be obtained by reengineering the toggle switch to express two different fluorescent proteins: one controlled by the same promoter as LacI and the other controlled by the same promoter as λcI. With an auto-sampler and multiple frequency fluorescence detectors, these new constructs can then be automatically measured with a time resolution of less than ten minutes–even including many samples from independent colonies. Furthermore, background fluorescence levels can be calibrated out of the data using mutants lacking one or both of the reporter proteins, and extrinsic noise can be decreased using forward and side scatter information (or even flow cytometry images) to restrict measurements to cells with similar shapes, sizes, and densities. In turn, such a carefully constrained and identified quantitative model could be used to predict switching behavior under different environmental conditions as well as help direct the modification of the circuit to meet other synthetic design objectives.

Parameter	Units	Set 0: **Ref. Set**	**Relative Parameter Values**				
			Set 1: μ_{LacI}	Set 2: $\mu_{\{\mathrm{LacI},\lambda\mathrm{cI}\}}$	Set 3: f_{LacI}	Set 4: $f_{\{\mathrm{LacI},\lambda\mathrm{cI}\}}$	Set 5: Full **P**
$k_{\lambda\mathrm{cI}}^{(0)}$	s^{-1}	$\mathbf{6.8}\times10^{-5}$	3.1×10^{0}	5.3×10^{0}	8.9×10^{0}	3.2×10^{0}	1.0×10^{0}
$k_{\lambda\mathrm{cI}}^{(1)}$	s^{-1}	$\mathbf{1.6}\times10^{-2}$	6.3×10^{-1}	7.6×10^{-1}	5.0×10^{-1}	9.1×10^{-1}	1.0×10^{0}
$k_{\mathrm{LacI}}^{(0)}$	s^{-1}	$\mathbf{2.2}\times10^{-3}$	3.0×10^{-1}	9.0×10^{-1}	5.6×10^{-1}	8.0×10^{-1}	1.0×10^{0}
$k_{\mathrm{LacI}}^{(1)}$	s^{-1}	$\mathbf{1.7}\times10^{-2}$	2.8×10^{-1}	6.5×10^{-1}	5.7×10^{-1}	8.0×10^{-1}	1.0×10^{0}
δ_{LacI}	$\mathrm{N}^{-1}\mathrm{s}^{-1}$	$\mathbf{3.8}\times10^{-4}$	2.2×10^{-1}	6.7×10^{-1}	5.6×10^{-1}	8.0×10^{-1}	1.0×10^{0}
$\delta_{\lambda\mathrm{cI}}(0)$	$\mathrm{N}^{-1}\mathrm{s}^{-1}$	$\mathbf{3.8}\times10^{-4}$	2.9×10^{-1}	6.1×10^{-1}	6.7×10^{-1}	9.5×10^{-1}	1.0×10^{0}
$\delta_{\lambda\mathrm{cI}}(6)$	$\mathrm{N}^{-1}\mathrm{s}^{-1}$	$\mathbf{6.7}\times10^{-4}$	5.9×10^{-1}	7.2×10^{-1}	7.2×10^{-1}	9.7×10^{-1}	1.0×10^{0}
$\delta_{\lambda\mathrm{cI}}(6)$	$\mathrm{N}^{-1}\mathrm{s}^{-1}$	$\mathbf{1.5}\times10^{-3}$	1.0×10^{0}	1.2×10^{0}	5.9×10^{-1}	9.5×10^{-1}	1.0×10^{0}
α_{LacI}	$\mathrm{N}^{-\eta_{\mathrm{LacI}}}$	$\mathbf{6.1}\times10^{-3}$	1.2×10^{0}	2.1×10^{0}	7.1e-02	5.9×10^{-1}	1.0×10^{0}
$\alpha_{\lambda\mathrm{cI}}$	$\mathrm{N}^{-\eta_{\lambda\mathrm{cI}}}$	$\mathbf{2.6}\times10^{-3}$	1.9×10^{-1}	5.4×10^{-1}	4.0×10^{-3}	4.8×10^{-1}	1.0×10^{0}
η_{LacI}	$--$	$\mathbf{2.1}\times10^{0}$	1.2×10^{0}	9.0×10^{-1}	1.3×10^{0}	1.1×10^{0}	1.0×10^{0}
$\eta_{\lambda\mathrm{cI}}$	$--$	$\mathbf{3.0}\times10^{0}$	1.1×10^{0}	1.1×10^{0}	1.7×10^{0}	1.1×10^{0}	1.0×10^{0}

Table 1: Parameter Sets for the stochastic toggle model, which have been identified from different data sets at a fixed time of five hours. For parameter sets 1-5, the table lists the relative parameter values. The five different data sets correspond to (1) the mean level of LacI, (2) the mean levels of LacI and λcI, (3) the marginal distribution of LacI, (4) the marginal distributions of LacI and λcI, and (5) the full joint distribution of LacI and λcI. The model responses for each of these parameter sets are shown in Figs. 2-5.

Parameter	Units	Set 0: **Ref. Set**	**Relative Parameter Values**				
			Set 6: μ_{LacI}	Set 7: $\mu_{\{\mathrm{LacI},\lambda\mathrm{cI}\}}$	Set 8: f_{LacI}	Set 9: $f_{\{\mathrm{LacI},\lambda\mathrm{cI}\}}$	Set 10: Full **P**
$k_{\lambda\mathrm{cI}}^{(0)}$	s^{-1}	$\mathbf{6.8}\times10^{-5}$	7.7×10^{-7}	1.0×10^{0}	9.9×10^{-1}	1.0×10^{0}	1.0×10^{0}
$k_{\lambda\mathrm{cI}}^{(1)}$	s^{-1}	$\mathbf{1.6}\times10^{-2}$	1.3×10^{0}	9.7×10^{-1}	1.0×10^{0}	1.0×10^{0}	1.0×10^{0}
$k_{\mathrm{LacI}}^{(0)}$	s^{-1}	$\mathbf{2.2}\times10^{-3}$	1.5×10^{0}	1.1×10^{0}	1.0×10^{0}	1.0×10^{0}	1.0×10^{0}
$k_{\mathrm{LacI}}^{(1)}$	s^{-1}	$\mathbf{1.7}\times10^{-2}$	9.6×10^{-1}	9.6×10^{-1}	1.0×10^{0}	1.0×10^{0}	1.0×10^{0}
δ_{LacI}	$\mathrm{N}^{-1}\mathrm{s}^{-1}$	$\mathbf{3.8}\times10^{-4}$	1.0×10^{0}	9.8×10^{-1}	1.0×10^{0}	1.0×10^{0}	1.0×10^{0}
$\delta_{\lambda\mathrm{cI}}(0)$	$\mathrm{N}^{-1}\mathrm{s}^{-1}$	$\mathbf{3.8}\times10^{-4}$	5.1×10^{-1}	9.7×10^{-1}	1.0×10^{0}	1.0×10^{0}	1.0×10^{0}
$\delta_{\lambda\mathrm{cI}}(6)$	$\mathrm{N}^{-1}\mathrm{s}^{-1}$	$\mathbf{6.7}\times10^{-4}$	7.5×10^{-1}	9.9×10^{-1}	1.0×10^{0}	1.0×10^{0}	1.0×10^{0}
$\delta_{\lambda\mathrm{cI}}(12)$	$\mathrm{N}^{-1}\mathrm{s}^{-1}$	$\mathbf{1.5}\times10^{-3}$	8.3×10^{-1}	1.0×10^{0}	1.0×10^{0}	1.0×10^{0}	1.0×10^{0}
α_{LacI}	$\mathrm{N}^{-\eta_{\mathrm{LacI}}}$	$\mathbf{6.1}\times10^{-3}$	5.7×10^{-1}	9.2×10^{-1}	1.0×10^{0}	1.0×10^{0}	1.0×10^{0}
$\alpha_{\lambda\mathrm{cI}}$	$\mathrm{N}^{-\eta_{\lambda\mathrm{cI}}}$	$\mathbf{2.6}\times10^{-3}$	5.2×10^{-1}	5.7×10^{-1}	1.0×10^{0}	1.0×10^{0}	1.0×10^{0}
η_{LacI}	$--$	$\mathbf{2.1}\times10^{0}$	1.2×10^{0}	1.0×10^{0}	1.0×10^{0}	1.0×10^{0}	1.0×10^{0}
$\eta_{\lambda\mathrm{cI}}$	$--$	$\mathbf{3.0}\times10^{0}$	1.0×10^{0}	1.1×10^{0}	1.0×10^{0}	1.0×10^{0}	1.0×10^{0}

Table 2: Parameter Sets for the stochastic toggle model, which have been identified from different data sets at all of five different measurement times: $\{1,2,3,4,5\}$ hours. The five different data sets are as listed in Table 1. For parameter sets 6-10, the table lists the relative parameter values.

Parameter Set #	**One Time Point at 5hr**					**All Time Points at {1,2,3,4,5}hr**				
	μ_{LacI}	$\mu_{\{\text{LacI},\lambda\text{cI}\}}$	f_{LacI}	$f_{\{\text{LacI},\lambda\text{cI}\}}$	Full **P**	μ_{LacI}	$\mu_{\{\text{LacI},\lambda\text{cI}\}}$	f_{LacI}	$f_{\{\text{LacI},\lambda\text{cI}\}}$	Full **P**
1	$< 10^{-3}$	1.42	0.78	2.4	2.2	3.2	8.1	10	16	15
2	$< 10^{-3}$	$< 10^{-3}$	0.91	1.8	1.3	0.59	1.1	5.5	9.4	6.7
3	0.004	1.4	0.015	1.3	1.4	0.81	8.8	3.9	11.4	8.5
4	0.007	0.017	0.024	0.043	0.072	0.61	1.3	1.4	2.1	1.6
5	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	0.002	0.006	0.007	0.012	0.010
6	0.002	1.7	0.68	2.7	2.3	0.023	7.6	2.8	11.2	11.2
7	0.001	0.006	0.18	0.28	0.19	0.011	0.046	0.72	1.1	0.78
8	$< 10^{-3}$	0.001	$< 10^{-3}$	0.001	0.001	$< 10^{-3}$	0.005	$< 10^{-3}$	0.005	0.008
9	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
10	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$

Table 3: The goodness of fit for each parameter set and each subset of considered data. The parameter sets correspond to the ten different fits found by using different amounts of the simulated data (see Tables 1 and 2). The different columns show how well each parameter set fits different aspects of the simulated data–the metric used is the one norm difference between the model with the true parameter set and the identified parameter set. For the mean values, the metric refers to relative difference (i.e., $d = |\mu_{model} - \mu_{data}|/\mu_{data}$).

References

[1] T. Gardner, C. Cantor, and J. Collins, "Construction of a genetic toggle switch in escherichia coli," *Nature*, vol. 403, pp. 339–242, 2000.

[2] H. Kobayashi, M. Kaern, M. Araki, K. Chung, T. Gardner, C. Cantor, and J. Collins, "Programmable cells: Interfacing natural and engineered gene networks," *PNAS*, vol. 101, pp. 8414–8419, June 2004.

[3] M. McAdams and A. Arkin, "Its a noisy business!," *Tren. Gen.*, vol. 15, no. 2, pp. 65–69, 1999.

[4] M. Elowitz, A. Levine, E. Siggia, and P. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, pp. 1183–1186, 2002.

[5] M. Thattai and A. van Oudenaarden, "Intrinsic noise in gene regulatory networks," *Proc. Natl. Acad. Sci.*, vol. 98, pp. 8614–8619, 2001.

[6] J. Hasty, J. Pradines, M. Dolnik, and J. Collins, "Noise-based switches and amplifiers for gene expression," *PNAS*, vol. 97, pp. 2075–2080, 2000.

[7] E. Ozbudak, M. Thattai, I. Kurtser, A. Grossman, and A. van Oudenaarden, "Regulation of noise in the expression of a single gene," *Nature Genetics*, vol. 31, pp. 69–73, 2002.

[8] N. Federoff and W. Fontana, "Small numbers of big molecules," *Science*, vol. 297, no. 5584, pp. 1129–1131, 2002.

[9] T. Kepler and T. Elston, "Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations.," *Biophys. J.*, vol. 81, pp. 3116–3136, 2001.

[10] J. Paulsson, O. Berg, and M. Ehrenberg, "Stochastic focusing: Fluctuation-enhanced sensitivity of intracellular regulation," *PNAS*, vol. 97, no. 13, pp. 7148–7153, 2000.

[11] H. Li, Z. Hou, and H. Xin, "Internal noise stochastic resonance for intracellular calcium oscillations in a cell system," *Phys. Rev. E*, vol. 71, no. 061916, 2005.

[12] A. Arkin, J. Ross, and M. H., "Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected escherichia coli cells," *Genetics*, vol. 149, pp. 1633–1648, 1998.

[13] D. Wolf and A. Arkin, "Fifteen minutes of fim: Control of type 1 pili expression in e. coli," *OMICS: A Journal of Integrative Biology*, vol. 6, pp. 91–114, Jan. 2002.

[14] B. Munsky, A. Hernday, D. Low, and M. Khammash, "Stochastic modeling of the pap-pili epigenetic switch," *Proc. FOSBE*, pp. 145–148, August 2005.

[15] T. Tian and K. Burrage, "Stochastic models for regulatory networks of the genetic toggle switch," *PNAS*, vol. 103, pp. 8372–8377, May 2006.

[16] Y. Dublanche, K. Michalodimitrakis, N. Kummerer, M. Foglierini, and L. Serrano, "Noise in transcription negative feedback loops: simulation and experimental analysis," *Molecular Systems Biology*, vol. 2, no. 41, 2006.

[17] A. Raj and A. van Oudenaarden, "Single-molecule approaches to stochastic gene expression," *Annual Review of Biophysics*, vol. 38, pp. 255–270, 2009.

[18] H. Shapiro, *Practical Flow Cytometry*. Wiley-Liss, 4 ed., 2003.

[19] A. Raj, P. van den Bogaard, S. Rifkin, A. van Oudenaarden, and S. Tyagi, "Imaging individual mrna molecules using multiple singly labeled probes," *Nature Methods*, vol. 5, pp. 877–887, 2008.

[20] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *J. Phys. Chem.*, vol. 81, pp. 2340–2360, May 1977.

[21] D. T. Gillespie, "The chemical langevin equation," *J. Chem. Phys.*, vol. 113, pp. 297–306, Jul. 2000.

[22] R. Allen, P. Warren, and P. Rein ten Wolde, "Sampling rare switching events in biochemical networks," *Phys. Rev. Lett.*, vol. 94, Jan. 2005.

[23] A. Warmflash, P. Bhimalapuram, and A. Dinner, "Umbrella sampling for nonequilibrium processes," *J. Chem. Phys.*, vol. 127, no. 154112, 2007.

[24] D. T. Gillespie, "Approximate accelerated stochastic simulation of chemically reacting systems," *J. Chem. Phys.*, vol. 115, pp. 1716–1733, Jul. 2001.

[25] C. V. Rao and A. P. Arkin, "Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the gillespie algorithm," *J. Chem. Phys.*, vol. 118, pp. 4999–5010, Mar. 2003.

[26] Y. Cao, D. Gillespie, and L. Petzold, "The slow-scale stochastic simulation algorithm," *J. Chem. Phys.*, vol. 122, Jan. 2005.

[27] N. van Kampen, *Stochastic Processes in Physics and Chemistry*. Elsevier, 3 ed., 2007.

[28] J. Elf and M. Ehrenberg, "Fast evaluations of fluctuations in biochemical networks with the linear noise approximation," *Genome Research*, vol. 13, pp. 2475–2484, 2003.

[29] I. Nasell, "An extension of the moment closure method," *Theoretical Population Biology*, vol. 64, pp. 233–239, 2003.

[30] C. Gmez-Uribe and G. Verghese, "Mass fluctuation kinetics: Capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations," *JCP*, vol. 126, Jan. 2007.

[31] A. Singh and J. Hespanha, "A derivative matching approach to moment closure for the stochastic logistic model," *Bulletin of Mathematical Biology*, vol. 69, pp. 1909–1925, 2007.

[32] N. Sinitsyn, N. Hengartner, and I. Nemenman, "Adiabatic coarse-graining and simulations of stochastic biochemical networks," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 106, no. 26, pp. 10546–10551, 2009.

[33] A. Walczak, A. Mugler, and C. Wiggins, "A stochastic spectral analysis of transcriptional regulatory cascades," *Proc. Nat. Acad. Sci.*, vol. 106, no. 16, pp. 6529–6534, 2009.

[34] B. Munsky and M. Khammash, "The finite state projection algorithm for the solution of the chemical master equation," *J. Chem. Phys.*, vol. 124, no. 044104, 2006.

[35] K. Burrage, M. Hegland, S. Macnamara, and R. Sidje, "A krylov-based finite state projection algorithm for solving the chemical master equation arising in the discrete modelling of biological systems," *Proc. of The A.A.Markov 150th Anniversary Meeting*, pp. 21–37, 2006.

[36] B. Munsky and M. Khammash, "The finite state projection approach for the analysis of stochastic noise in gene networks," *IEEE Trans. Automat. Contr./IEEE Trans. Circuits and Systems: Part 1*, vol. 52, pp. 201–214, Jan. 2008.

[37] B. Munsky, *The finite State projection Approach for the Solution of the Chemical Master Equation and its Application to Stochastic Gene Regulatory Networks*. PhD thesis, Univ. of California at Santa Barbara, Santa Barbara, June 2008.

[38] T. Cagatay, M. Turcotte, M. Elowitz, J. Garcia-Ojalvo, and G. Suel, "Architecture-dependent noise discriminates functionally analogous differentiation circuits," *Cell*, vol. 139, no. 3, pp. 512–522, 2009.

[39] A. Warmflash and A. Dinner, "Signatures of combinatorial regulation in intrinsic biological noise," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 45, pp. 17262–17267, 2008.

[40] M. Dunlop, R. Cox III, J. Levine, R. Murray, and M. Elowitz, "Regulatory activity revealed by dynamic correlations in gene expression noise," *Nature Genetics*, vol. 40, pp. 1493–1498, 2008.

[41] W. de Ronde, B. Daniels, A. Mugler, N. Sinitsyn, and I. Nemenman, "Mesoscopic statistical properties of multistep enzyme-mediated reactions," *IET Syst. Biol.*, vol. 3, no. 5, pp. 429–437, 2009.

[42] B. Munsky, B. Trinh, and M. Khammash, "Listening to the noise: random fluctuations reveal gene network parameters," *Molecular Systems Biology*, vol. 5, no. 318, 2009.

[43] D. Thorsley and E. Klavins, "Approximating stochastic biochemical processes with wasserstein pseudometrics," *IET Systems Biology*, vol. 4, no. 3, pp. 193–211, 2010.

[44] P. Warren and P. Rein ten Wolde, "Chemical models of genetic toggle switches," *J. Phys. Chem. B*, vol. 109, pp. 6812–6823, 2005.

[45] A. Lipshtat, A. Loinger, N. Balaban, and O. Biham, "Genetic toggle switch without cooperative binding," *Phys. Rev. Lett.*, vol. 96, no. 188101, 2006.

[46] B. Munsky and M. Khammash, "A multiple time interval finite state projection algorithm for the solution to the chemical master equation," *J. Comp. Phys.*, vol. 226, no. 1, pp. 818–835, 2007.

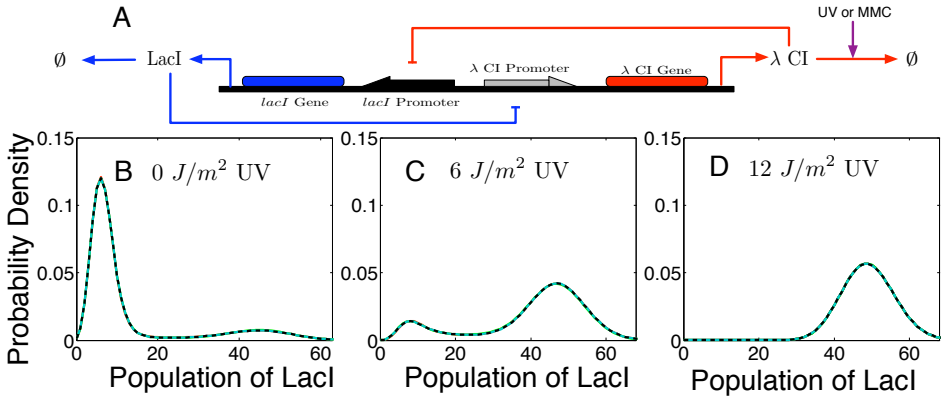

Figure 1: (A) Basic schematic of the toggle model comprised of two inhibitors: λcI inhibits the production of LacI and vice-versa. In the model, the synthesis rates of λcI and LacI are non-linear functions of their counterparts. Environmental influences (ultraviolet radiation) increase the degradation rate of λcI and affect the tradeoff between the two regulators. (B-D) Marginal probability distributions of LacI as simulated with parameters sets #0 and #3 from Table 1. The degradation parameter δ_u is the only parameter that changes between the panels B-D.

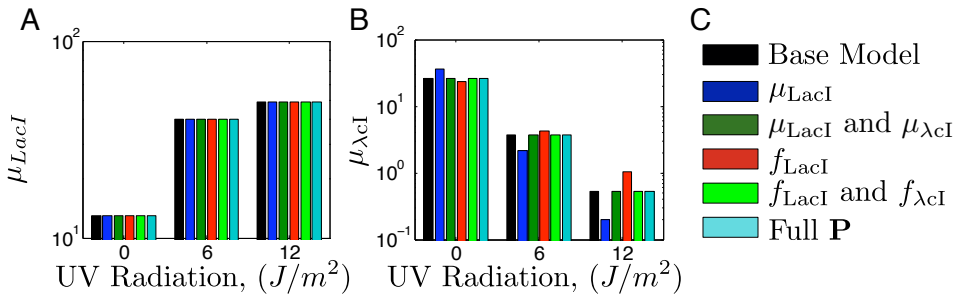

Figure 2: Mean levels of (A) LacI and (B) λcI for the toggle model, 5 hours after induction with 0, 6, and 12 J/m^2 ultraviolet radiation. (C) Five different parameter sets have been used corresponding to the data types that are fully captured by the given parameter sets listed in Table 1. For example, the blue bars (color online) correspond to parameter set #1, which matches the mean LacI level at t=5hr. Similarly, the red bars correspond to parameter set #3, which matches the LacI marginal distributions at at t=5hr.

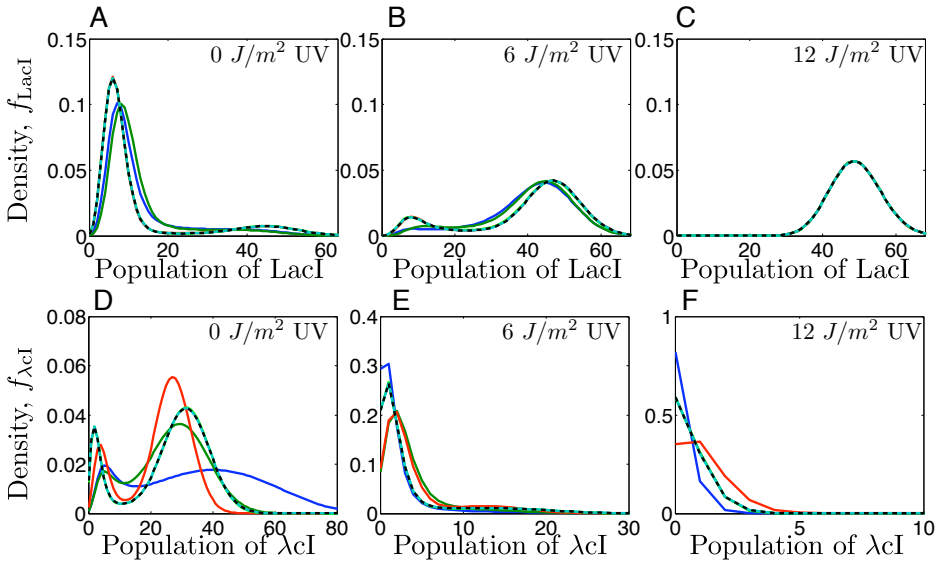

Figure 3: Marginal probability densities for (A-C) LacI and (D-F) λcI for the toggle model, at 5 hours after induction with 0, 6, and 12 J/m^2 ultraviolet radiation. (B) Five different parameter sets have been used corresponding to the data types that are fully captured by the given parameter sets listed in Table 1 (see Fig. 2C for legend.)

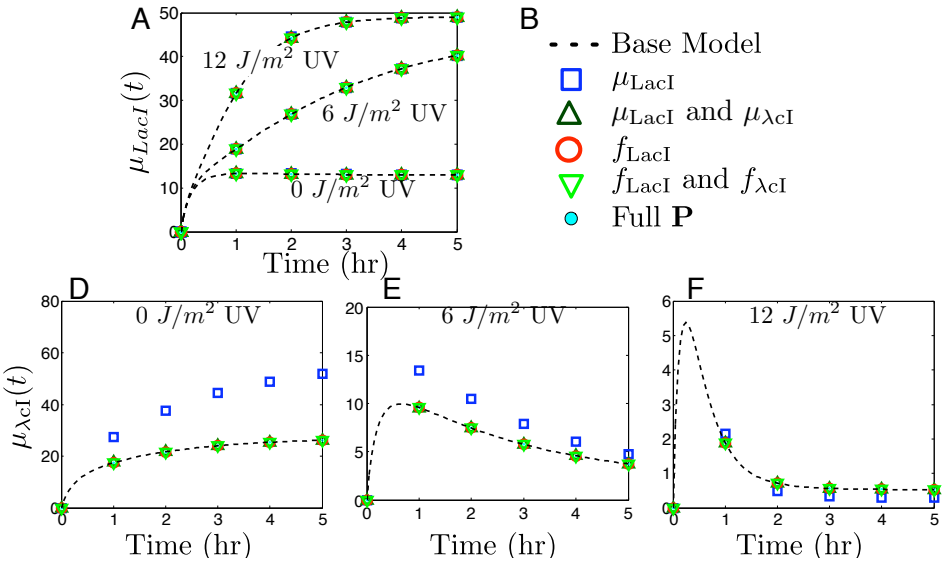

Figure 4: Mean levels of (A) LacI and (D-E) λcI for the toggle model, as functions of time after induction with 0, 6, and 12 J/m^2 ultraviolet radiation. (B) Five different parameter sets have been used corresponding to the data types that are fully captured by the given parameter sets listed in Table 2 (see also caption of Fig. 2.)